

Editorial

Vision, action and language unified through embodiment

Daniele Caligiore¹, Martin H. Fischer²

*¹Laboratory of Computational Embodied Neuroscience,
Istituto di Scienze e Tecnologie della Cognizione,
Consiglio Nazionale delle Ricerche (LOCEN-ISTC-CNR), Roma, Italy*

²Division of Cognitive Sciences, University of Potsdam, Potsdam, Germany

Corresponding author:

Dr. Daniele Caligiore

e-mail address: daniele.caligiore@istc.cnr.it

Running title: Vision, action and language unified through embodiment

Increasing evidence shows that vision, action and language should not be regarded as a set of disembodied processes. Instead, they form a closely *integrated* and highly dynamic system that is attuned to the constraints of its bodily implementation as well as to the constraints coming from the world with which this body interacts. One consequence of such embodiment of cognition is that seeing an object, even when there is no intention to handle it, activates plans for actions directed towards it (e.g., Tucker & Ellis, 1998; 2001; Fischer & Dahl, 2007). Using object names induces similar action planning effects as seeing the objects themselves (Tucker & Ellis, 2004; Borghi, Glenberg & Kaschak, 2004). Depending on linguistic context, different object features can be activated for action planning, as indicated by facilitated manual responses or “affordance effects” (e.g., Borghi, 2004; Glenberg & Robertson, 2000; Zwaan, 2004). Similarly, different action intentions direct attention differently to object features for processing (e.g., Bekkering & Neggers, 2002; Fischer & Hoellen, 2004; Symes, Tucker, Ellis, Vainio, & Ottoboni, 2008). Eye movements during visually guided actions shed further light on the close relationship between vision, action and language (Land & Furneaux, 1997; Johansson, Westling, Backstrom & Flanagan, 2001). For example when humans interact with objects, their eyes move ahead of their hands to support the on-line control of grasping (e.g., Bekkering & Neggers, 2002).

These behavioural results are supported by brain imaging studies of object affordances in humans (e.g., Grèzes, Tucker, Armony, Ellis & Passingham, 2003) and single cell recordings in monkeys (e.g., Sakata, Taira, Mine & Murata, 1992; Fadiga, Fogassi, Gallese & Rizzolatti, 2000). Together, these behavioural and neuroscientific studies have recently begun to inform computational models of embodied cognition. For example, Tsiotas, Borghi and Parisi (2005) devised an artificial life simulation to give an evolutionary account of some affordance effects, and Caligiore, Borghi, Parisi, and Baldassarre (2010) proposed a computational model to account for several affordance-related effects in grasping, reaching, and language. The neuroscientific constraints implemented in the design of the model allow its authors to investigate the neural mechanisms underlying

affordance selection and control. The present special issue brings together recent developments at the intersection between behavioral, neuroscientific, and computational approaches to embodied cognition.

Strong support for the close link between vision, action and language comes from studies which highlight how language processing and comprehension make use of neural systems ordinarily used for perception and action (Lakoff, 1987; Zwaan, 2004; Barsalou, 1999; Glenberg & Robertson, 1999; Gallese, 2008; Glenberg, 2008). For example, when humans process the word “cup” they seem to reenact (and therefore internally simulate) many of the perceptual, motor and affective representations related to a cup (Barsalou, 1999). In a similar way sentences and abstract words are understood by creating a simulation of the actions underlying them (Glenberg and Kaschak, 2002; see also Borghi & Cimatti, 2009, for a new formulation of embodiment of abstract words which includes social aspects). Moreover, when hearing a verbal description of a visually available scene humans tend to look at objects that are about to be mentioned (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), indicating rapid and predictive comprehension that is tightly linked to action. Several computational models have therefore implemented simple learning mechanisms (such as Hebbian rules) to create associations between patterns of active neurons representing the phonological aspects of words and internal simulations (i.e., representations of object features involved in perception and action, cf. Jeannerod, 2007; Mayor & Plunkett, 2010; Caligiore, Borghi, Parisi, Baldassarre, 2010; Li, Farkas, & MacWhinney, 2004).

Grounded cognition theories have found a neurophysiological basis in the recent discovery, in monkeys as well as in humans, of two kinds of visuomotor neurons: canonical and mirror neurons (Rizzolatti & Craighero, 2004; Buccino et al, 2004). Canonical neurons discharge to the visual presentation of objects that can be grasped with a specific type of prehension (object directed action), motorically coded by these neurons even when a grasping movement is not required. Mirror neurons, instead, fire when the monkey makes a goal-directed action and also when it observes another monkey or an experimenter performing the same or a similar action. Recent studies, mainly based on brain imaging techniques, indicate

the existence of both canonical and mirror neurons in humans (Buccino et al., 2001; Grèzes, Tucker, Armony, Ellis, & Passingham, 2003; Johnson-Frey, Maloof, Newman-Norlund, Farrer, Inati, & Grafton, 2003; Fadiga et al., 2006). Using fMRI it has been shown that in humans the observation of both object directed actions and mimed actions leads to activation of different regions in the premotor cortex, including the Broca's region (Buccino et al., 2001) and the parietal cortex (Fogassi, Ferrari, Gesierich, Rozzi, Chersi, & Rizzolatti, 2005). The relationship between canonical and mirror neurons and their roles in different cognitive functions, including language processing, has to be better investigated (see Thill et al., submitted, for an up-to-date review on these topics).

In an influential paper, Rizzolatti and Arbib (1998) proposed that the matching process embodied by mirror neurons represents the basic mechanism from which language evolved. In the last decade this claim has been strongly supported by a series of experimental studies (for reviews see Pulvermüller, 2005; Willems & Hagoort, 2007). First, in an event-related fMRI study, the silent reading of words referring to face, arm or leg actions activated premotor-motor areas related to the word meanings (Hauk et al 2004). An MEG study showed that reading action verbs activates motor and premotor cortices both rapidly and in a somatotopic fashion (Pulvermüller, 2005), thus suggesting that motor activation is inherent to lexical processing. In a further fMRI study, listening to sentences expressing mouth, hand and foot actions produced activation of effector-congruent sectors of the premotor cortex (Tettamanti et al., 2005). Interestingly, these distinct sectors coincide, albeit only approximately, with those active during the observation of hand, mouth and foot actions (Buccino et al., 2001).

These data support the notion that the mirror neuron system is involved not only in understanding visually presented actions, but also in coding acoustically presented action-related sentences. Several studies showed that similar mechanisms of *motor resonance* are active when we understand hand and mouth actions including speech production. First, grasping movements influence syllable pronunciation when executed (Gentilucci, 2001) as well as when merely observed (Gentilucci, 2003). Secondly, both listening and observing speech movements causes an increase of motor evoked potentials recorded from tongue and lip

muscles (Watkins, Strafella, & Paus, 2003; Pulvermüller & Fadiga, 2010). Finally, evidence for a link between gesturing and the speech system also comes from clinical studies: Hanlon, Brown and Gerstman (1990) showed that aphasic patients' object naming benefits from pointing with the right hand to the referents.

Investigations into the integration of vision, action and language have greatly benefited from the use of computational models. The linguistic abilities of an artificial agent whose behavior is established by a computational model is strictly dependent on, and grounded in, other perceptual and motor skills (MacWhinney, 1998; Cangelosi & Riga, 2006; Cangelosi & Parisi 2002). Such a grounded and embodied approach to language design is consistent with the theories of the grounding of language discussed above. In these models there exists an intrinsic link between the communication symbols (words) used by the agent and its own cognitive representations (meanings) of the perceptual and sensorimotor interaction with the external world (referents) (Steels & Vogt, 1997; Steels, 2003; Yoon, Heinke & Humphreys, 2002). Cangelosi, Hourdakis, and Tikhonoff (2006) proposed a neural network model in a robotic set-up as a model of language acquisition. The authors show how a robot can acquire new concepts of actions via linguistic instructions. Moreover, the associative mechanisms involving words and categorical representations of objects are used to transfer the compositionality properties of language to sensorimotor representations. In the same line Chersi, Thill, Ziemke and Borghi (2010) recently used a computational model to show how sentence processing might involve similar chaining mechanisms as does action sequence organization (Fogassi, Ferrari, Gesierich, Rozzi, Chersi, & Rizzolatti, 2005). Many other embodied modeling issues remain to be resolved (Pezzulo, Barsalou, Cangelosi, Fischer, Spivey, & McRae, 2011).

The various studies mentioned above investigate the integration of vision, action, and language through embodiment from rather different perspectives. We have discussed results from behavioral experiments as well as from neuroscientific and computational modeling. Unfortunately, despite the converging results obtained across these disciplines there is currently very *little proper discussion and exchange of views* among the experts that use these different perspectives. We believe that this kind of multi-methodological and multidisciplinary discussion is

a useful and necessary step to share the advantages of each approach and to achieve a cumulative understanding of the neural mechanisms the brain uses to deal with the integration of vision, action and language into embodied behaviour. Instead of stressing the differences between the different approaches, it could be productive to focus on their overlapping traits and on their enormous potential for cross-fertilization. *The aim of this special issue* is therefore to direct the attention of scientists who work with different approaches towards an inter-disciplinary discussion about vision, action and language unified through embodiment.

CONTENT OF THE SPECIAL ISSUE

This issue of *Psychological Research* has a clear focus: understanding how the results from different scientific approaches could be shared to energize the empirical and theoretical discussion on the integration of vision, action and language by embodiment. The contributions to this special issue cover a wide range of methodologies, from psychophysics to computational modeling, from classical behavioural methods to neuropsychological and brain imaging approaches. The majority of contributions goes beyond the popular reaction time methodology and uses movement-related performance to strengthen the case for an embodiment of concept activation.

Some themes explored in the special issue regard: What is the organization (intended at several levels such as brain, functional, computational) needed to support integration of vision, action and language? How does the timing in language processing influence the embodiment of language representations? What is the relationship between canonical and mirror neurons in action and language organization? What is the social influence on affordances perception and organization? Is the motor system involved in understanding of concrete nouns, as it is for concrete verbs? The range of perspectives of the papers proposed in this special issue offers psychological, neuroscientific and computational modeling evidence in the investigation of these questions.

Caligiore, Borghi, Parisi, Ellis, Cangelosi, and Baldassarre (2012), propose an extended version of the TRoPICALS computational model (Caligiore, Borghi, Parisi, & Baldassarre, 2010) aimed at better understanding the mechanisms underlying *positive as well as negative compatibility effects* observed in

behavioural experiments. The model addresses the case of distractor objects which, although irrelevant for the agent's goals, activate affordances that have to be actively suppressed. The simulations fully replicate the findings reported in the literature. The authors further simulate damages to the model that are similar to those found in Parkinson's Disease in order to predict compatibility effects that might be found with these patients in future experiments.

De Vega, Moreno, and Castillo (2012) present two experiments that look at changes in motor compatibility effects during comprehension based on the *relative timing of the motor response to the processing of action-relevant language*. The authors show that at short Stimulus Onset Asynchrony, the traditional motor compatibility effect is reversed: participants are faster to respond when the direction of the action in the sentence mismatches the direction of the motor response that needs to be made. The work deals with a timely and important issue, and the data help to reconcile some differing results that have been reported about facilitating and interfering observations of motor compatibility effects.

Ellis, Swabey, Bridgeman, May, Tucker, and Hyne (2012) report a behavioral study to investigate the *interaction of the mirror neuron system and the canonical neuron system* when humans observe other agents acting on objects irrespective of their goals. They make a case for regarding them as different aspects of a common system for orchestrating the actions of agents.

Gianelli, Scorolli, and Borghi (2012) present an empirical study to investigate the effects of *social influences* on kinematic features of a reaching and grasping movements. They recorded reaching and grasping movements in the presence of a second person which could be either a friend or no friend. The authors demonstrate that the social relationship between a performer and a second person affected kinematic features of the task. Moreover, speaking sentences related to the reaching and grasping task had an effect depending on whether "I" or "you" was used as a pronoun. These results point in the direction of social motor control as a novel field of embodiment research.

Iizuka, Marocco, Ando, and Maeda (2012) present an empirical analysis of how a communication system emerges spontaneously between two interacting individuals in the absence of a specifically predefined communication channel. Participants tried to communicate to each other the identity of viewed objects by

sliding fingers on a signaling device. The emerging communication patterns suggested gradual emergence of turn taking, association between behaviour and perceptual categories, and the acquisition of novel meanings. These observations investigate the foundations of our sociality.

Marino, Gough, Gallese, Riggio, and Buccino (2012) offer new empirical evidence of embodied meaning associated to action-related nouns rather than verbs. The work addresses the crucial open question of whether the *motor system* is *involved* during the understanding of *concrete nouns*, as it is for *concrete verbs*. The results are discussed in terms of motor processes in the left brain hemisphere associated with action nouns.

Weiner and Grill-Spector (2012) summarize the results of two recently published studies (Weiner & Grill-Spector, 2010; 2011) that investigated the distribution of face and limb selectivity in human visual cortex. They propose a new three-stream model of high-level visual cortex which includes ventral, lateral and dorsal areas where multimodal processing related to vision, action and language might converge. Just as the other contributions to this special issue, this programmatic proposal sets a framework for a much needed dialogue between disciplines.

TOWARDS A COMMON FRAMEWORK TO STUDY THE EMBODIMENT OF VISION, ACTION AND LANGUAGE

The accumulation of evidence in favour of embodied cognition, which comes from such different disciplines as psychology, neuroscience and robotics, confirms the importance of the topic of this special issue for the wider scientific community. However, the multi-disciplinary and multi-methodological nature of the available data raises an important question: Is it possible to find a *common framework* to interpret and explain data deriving from such vastly different methods? This is a crucial point because such a common framework could support cross-fertilization among different disciplines and, importantly, could help to discover general principles underlying the embodiment of vision, action and

language. However, considering a framework that is understandable by scientists with dramatically different backgrounds who often use different terminology to indicate the similar phenomena is not trivial (Hommel & Colzato, 2010).

In the last decade some valid attempts, mainly using computational approaches, have been proposed (Arbib & Lee, 2007; Garagnani, Wennekers, & Pulvermüller, 2008; O'Reilly, 1998; Rothkopf & Ballard, 2010). Arbib and colleagues designed several models, including the FARS model (Fagg & Arbib, 1998) and various incarnations of the MNS models (MNS: Oztop & Arbib, 2002; MNS2-I: Bonaiuto, Rosta, & Arbib, 2007; MNS2-II: Bonaiuto & Arbib, 2010), that might be conducive to the intended cross-disciplinary investigation the topic of this special issue. Two other proposals merit attention in this regard since they have started to *formalize some procedures to build cross-disciplinary frameworks* to investigate psychological and neuroscientific phenomena. These two methods are: the *Brain-Based Devices (BBDs)* approach (Fleischer & Edelman, 2009) and the *Computational Embodied Neuroscience (CEN)* approach (Caligiore, Borghi, Parisi, & Baldassarre, 2010; Mannella, Mirolli, & Baldassarre, 2010; cf. Prescott, Montes-Gonzalez, Gurney, Humphries, & Redgrave, 2006, for a similar but less principled approach).

The BBDs approach and the CEN method are similar in conception. The key features of models based on these two methods are: (a) a simulated brain whose anatomy and physiology is constrained by knowledge about real brains; (b) an embodied system which operates in a real environment; (c) the comparison with data from behavioral experiments; (e) the adaptive learning of the behavior. However, differently from BBDs, the CEN approach is also guided by the further and fundamental meta-constraint of *theoretical cumulativity*. This idea aims at producing general models that account for an increasing number of experiments, avoiding at the same time to build ad-hoc models which account for only specific single experiments. In this way it could be possible to isolate general principles

underlying the class of studied phenomena, thereby producing theoretical cumulativity.

To facilitate the integration among different perspectives and different methods it will also be crucial to design “system-level models”. This means that the main goal of the model should be to provide an operational hypothesis about the cerebral network of networks which underlies the investigated behavior. The system-level approach postulates that the different classes of behaviors are generated by the interplay of different subsets of components of the brain, rather than by specific components in isolation. In this way it will be possible to outline an integrated hypothesis about the system-level architectural and functioning brain mechanisms which might underlie the behaviour under investigation. For example, a system-level model might take into account both cortical (Rizzolatti & Arbib, 1998) and sub-cortical (Strick, Dum, & Fiez, 2009) mechanisms underlying the embodiment of language, or might facilitate the interpretation of brain imaging data (Friston, 2009). We hope that, in the future, designing theoretical and computational frameworks using multi-disciplinary approaches such as those proposed by the BBDs and CEN methods will help to provide a unified view of embodiment of vision, action, and language, highlighting all its challenging aspects and fostering further research into this exciting topic.

Acknowledgements

We would like to thank all the reviewers that with their work have greatly enhanced the quality of this special issue and Dr. Gianluca Baldassarre for his precious comments on this editorial. This work was supported by the EU funded Projects “IM CLeVeR – Intrinsically Motivated Cumulative Learning Versatile Robots”, contract no. FP7-IP – 231722, and “VALUE – Vision, Action, and Language Unified by Embodiment” , EPSRC Grant EP/F026471.

REFERENCES

- Arbib, M. A. (1997). From visual affordances in monkey parietal cortex to hippocampo-parietal interactions underlying rat navigation. *Philosophical Transactions of The Royal Society B Biological Sciences*, 352, 1429–1436.

- Arbib, M. A., & Lee, J. Y. (2007). Vision and Action in the Language-Ready Brain: From Mirror Neurons to SemRep. *Lecture Notes in Computer Science*, 4729, 104.
- Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Bekkering, H., & Neggers, S. F. W. (2002). Visual search is modulated by action intentions. *Psychological Science*, 13, 370–374.
- Bonaiuto, J., Rosta, E. & Arbib, M. A. (2007). Extending the mirror neuron system model, I. *Biological Cybernetics*, 96, 9–38.
- Bonaiuto, J. & Arbib, M. A. (2010). Extending the mirror neuron system model, II: what did I just do? A new role for mirror neurons. *Biological Cybernetics*, 102, 341–259.
- Borghi, A. M., Glenberg, A. M., & Kaschak, M. (2004). Putting words in perspective. *Memory & Cognition*, 32, 863–873.
- Borghi, A. M. (2004). Object concepts and action: extracting affordances from objects parts. *Acta Psychologica*, 115(1), 69–96.
- Borghi, A. M., & Cimatti, F. (2009). Words as tools and the problem of abstract words meanings. In N. Taatgen & H. van Rijn (eds.). *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2304-2309). Amsterdam: Cognitive Science Society.
- Buccino, G., Binkofski, F., & Riggio, L. (2004). The mirror neuron system and action recognition. *Brain and Language*, 89, 370–376.
- Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., & Freund, H.J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, 13, 400–404.
- Caligiore, D., Borghi, A. M., Parisi, D., & Baldassarre, G. (2010). TRoPICALS: A Computational Embodied Neuroscience Model of Compatibility Effects. *Psychological Review*, 117, 1188–1228.
- Caligiore, D., Borghi, A. M., Parisi, D., Ellis, R., Cangelosi, A., & Baldassarre, G. (2012). How affordances associated with a distractor object affect compatibility effects: A study with the computational model TRoPICALS. *Psychological Research*, this issue.
- Cangelosi, A., & Riga, T. (2006). An embodied model for sensorimotor grounding and grounding transfer: experiments with robots. *Cognitive Science*, 30, 673–689.
- Cangelosi, A., & Parisi, D. (2002). *Simulating the Evolution of Language*. In A. Cangelosi & D. Parisi, Eds. *Simulating the Evolution of Language* (p. xii + 356). Springer-Verlag.
- Cangelosi, A., Hourdakis, E., & Tikhonoff, V. (2006). Language Acquisition and Symbol Grounding Transfer with Neural Networks and Cognitive Robots. *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, 1576–1582.
- Chersi, F., Thill, S., Ziemke, T. & Borghi, A. M. (2010) Sentence processing: linking language to motor chains. *Frontiers in Neurobotics* 4(4), doi:10.3389/fnbot.2010.00004.
- Craigheo, L., Bello, A., Fadiga, L., Rizzolatti, G. (2002). Hand action preparation influences the responses to hand pictures. *Neuropsychologia*, 40, 492–502

- De Vega, M., Moreno, V., & Castillo, D. (2012). The comprehension of action-related sentences may cause interference rather than facilitation on matching actions. *Psychological Research*, this issue.
- Ellis, R., Swabey, D., Bridgeman, J., May, B., Tucker, M., & Hyne, A. (2012). Bodies and other visual objects: the dialectics of reaching toward objects. *Psychological Research*, this issue.
- Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G. (2000). Visuomotor neurons: ambiguity of the discharge or 'motor' perception? *International Journal of Psychophysiology*, 35, 165–177.
- Fadiga, L., Craighero, L., Fabbri Destro, M., Finos, L., Cotillon Williams, N., Smith, A. T., & Castiello, U. (2006). Language in shadow. *Social Neuroscience*, 1, 77–89.
- Fagg, A. H., & Arbib, M. A., (1998). Modeling parietal-premotor interaction in primate control of grasping. *Neural Networks*, 11, 1277–1303.
- Fischer, M. H., & Dahl, C. (2007). The time course of visuo-motor affordances. *Experimental Brain Research*, 176, 519–524.
- Fischer, M. H., & Hoellen, N. (2004). Space- and object-based attention depend on motor intention. *The Journal of General Psychology*, 13, 365–377.
- Fleischer, J. G. & Edelman, G. M. (2009). Brain-based devices: An embodied approach to linking nervous system structure and function to behavior. *IEEE Robotics & Automation Magazine*, 16, 33–41.
- Fodor, J. A., & Pylyshyn, Z.W. (1988) Connectionism and cognitive architecture: a critical analysis. *Cognition*, 28, 3–71.
- Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F. & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science*, 308, 662–667.
- Friston, K. (2009). Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLoS Biol*, Vol. 7 (No. 2). pp. e33.
- Gallese, V. (2008). Mirror neurons and the social nature of language: the neural exploitation hypothesis. *Social Neuroscience*, 3, 317–333.
- Garagnani, M., Wennekers, T., Pulvermüller, F. (2008). A neuroanatomically grounded Hebbian-learning model of attention-language interactions in the human brain. *European Journal of Neuroscience*, 27, 492–513.
- Gentilucci, M., Benuzzi, F., Gangitano, M., & Grimaldi, S. (2001). Grasp with hand and mouth: a kinematic study on healthy subjects. *Journal of Neurophysiology*, 86, 1685–1699.
- Gentilucci, M. (2003). Grasp observation influences speech production. *European Journal of Neuroscience*, 17, 179–184.
- Gianelli, C., Scorolli, C., & Borghi, A. M. (2012). Acting in perspective: The role of body and language as social tools. *Psychological Research*, this issue.
- Glenberg, A. M., Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory & Language*, 43, 379–401.
- Glenberg, A.M., & Robertson, D.A. (1999). Indexical understanding of instructions. *Discourse Processes*, 28, 1–26.

- Glenberg, A.M., Kaschak, M.P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9, 558-565.
- Glenberg A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 586–596.
- Grèzes, J., Tucker, M., Armony, J., Ellis, R., & Passingham, R. E. (2003). Objects automatically potentiate action: an fMRI study of implicit processing. *European Journal of Neuroscience*, 17, 2735–2740.
- Hanlon, R. E., Brown, J.W., & Gerstman, L.J. (1990). Enhancement of naming in nonfluent aphasia through gesture. *Brain and Language*, 38, 298–314.
- Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41, 301–307.
- Hommel, B., & Colzato, L. S. (2010). Games with(out) Frontiers: Towards an integrated science of human cognition. *Frontiers in Psychology*, 1:2.
- Iizuka, H., Marocco, D., Ando, H., & Maeda, T. (2012). An Experimental Study on Co-evolution of Categorical Perception and Communication System in Humans. *Psychological Research*, this issue.
- Jeannerod, M. (2007). *Motor Cognition: What actions tell the self*. Oxford: Oxford University Press.
- Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye-hand coordination in object manipulation. *Journal of Neuroscience*, 21, 6917–6932.
- Johnson-Frey, S. H., Maloof, F. R., Newman-Norlund, R., Farrer, C., Inati, S., & Grafton, S. T. (2003). Actions or hand-objects interactions? Human inferior frontal cortex and action observation. *Neuron*, 39, 1053–1058.
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences*, 352(1358), 1231–1239.
- Li, P., Farkas, I., & MacWhinney, B. (2004) Early lexical development in a self-organizing neural network. *Neural Networks* 17, 1345–1362.
- Marino, B., Gough, P. M., Gallese, V., Riggio, L., & Buccino, G. (2012). How the motor system handles nouns: A behavioural study. *Psychological Research*, this issue.
- Mannella, F., Mirolli, M., & Baldassarre, G. (2010). The interplay of Pavlovian and instrumental processes in devaluation experiments: A computational embodied neuroscience model tested with a simulated rat. In Tosh, C.R. and Ruxton, G.D. (eds.), *Modelling perception using artificial neural networks*. Cambridge: Cambridge University Press.
- Mayor, J., & Plunkett, K. (2010). A neurocomputational account of taxonomic responding and fast mapping in early word learning. *Psychological Review*, 117, 1–31.
- MacWhinney, B. (1998). Models of the emergence of language. *Annual Review of Psychology*, 49, 199–227.

- O'Reilly, R.C. (1998). Six Principles for Biologically-Based Computational Models of Cortical Cognition. *Trends in Cognitive Sciences*, 2, 455–462.
- Oztop, E. & Arbib, M. A. (2002). Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics*, 87, 116–140.
- Pezzulo, G., Barsalou, L.W., Cangelosi, A., Fischer, M.H., Spivey, M., & McRae, K. (2011). The Mechanics of Embodiment: A Dialogue on Embodiment and Computational Modeling. *Frontiers in Psychology*, doi: 10.3389/fpsyg.2011.00005.
- Prescott, T. J., Montes-Gonzalez, F., Gurney, K., Humphries, M. D., & Redgrave, P. (2006). A robot model of the basal ganglia: Behavior and intrinsic processing. *Neural Networks*, 19, 31–61.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Review Neuroscience*, 6, 576–582.
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11, 351–60.
- Pylyshyn, Z.W. (1973) What the mind's eye tells the mind's brain: a critique of mental imagery. *Psychological Bulletin*, 80, 1–24.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169-192.
- Rizzolatti, G. & Arbib, M.A. (1998). Language within our grasp. *Trends in Neurosciences*, 21, 188–194.
- Rothkopf, C. A., & Ballard, D. H. (2010). Credit assignment in multiple goal embodied visuomotor behavior, *Frontiers in Psychology*.
- Sakata, H., Taira, M., Mine, S., & Murata, A. (1992). Hand-movement-related neurons of the posterior parietal cortex of the monkey: Their role in the visual guidance of hand movements. In R. Caminiti, P. B. Johnson, Y. Burnod, *Control of arm movement in space: Neurophysiological and computational approaches* (pp. 185–198). Heidelberg: Springer.
- Strick, P. L., Dum, R. P., Fiez, J. A. (2009). Cerebellum and non-motor function. *Annual Review of Neuroscience*, 32, 413–434.
- Symes, E., Tucker, M., Ellis, R., Vainio, L., & Ottoboni, G. (2008). Grasp preparation improves change detection for congruent objects. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 854–71.
- Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M. & Sedivy, J.E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P. Fazio, F., Rizzolatti, G., Cappa, S. F. & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, 17, 273–281.
- Thill, S., Caligiore, D., Borghi, A. M., Ziemke, T., & Baldassarre, G. (2012). Integrating Theories and Computational Models of Affordance Control and Mirror Neurons: A Review and a Model Design. Submitted to *Neuroscience & Biobehavioral Reviews*.

- Tsiotas, G., Borghi, A. M., & Parisi, D. (2005). Objects and affordances: An Artificial Life simulation. In B. Bara, L. Barsalou, B. Bucciarelli (Eds.), COGSCI2005. XXVII Annual Conference of the Cognitive Science Society (pp. 2212–2217). Mahwah, N. J.: Lawrence Erlbaum Associates.
- Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 830–846.
- Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition*, 8, 769–800.
- Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica*, 116, 185–203.
- Watkins, K.E., Strafella, A.P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41, 989–994.
- Weiner, K. S., & Grill-Spector, K. (2012). Neural representations of faces and limbs neighbor in human high-level visual cortex: Evidence for a new organization principle. *Psychological Research*, this issue.
- Willems, R.M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture and action: a review. *Brain and Language*, 101, 278–289.
- Zwaan, R.A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B.H. Ross (ed.), *Psychology of learning and motivation*, (vol.44, pp.35-62). New York: Academic.